

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:49:35 ; Search time 140.11 Seconds  
(without alignments)  
8.074 Million cell updates/sec

Title: US-09-331-631A-3\_COPY\_186\_248

Perfect score: 353  
Sequence: 1 KRDPQQRHYEDCRRRCQQE.....LINDQCGSGRYEGEEKQS 63

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118.5	33.6	566	1	US-07-955-905A-2
2	118.5	33.6	566	1	US-07-955-905A-22
3	111.5	31.6	587	1	US-07-955-905A-23
4	71.5	20.3	1898	1	US-08-056-200-94
5	71.5	20.3	1898	2	US-08-800-644-94
6	67	19.0	1162	2	US-08-728-323A-2
7	66	18.7	614	4	PCF-US95-03236-21
8	65.5	18.6	816	2	US-08-267-803B-9
9	65	18.4	571	1	US-07-955-905A-25
10	65	18.4	2476	2	US-08-276-967-2
11	64.5	18.3	34	2	US-08-822-561-1
12	64.5	18.3	361	1	US-08-415-751-4
13	63	17.8	365	1	US-08-437-027-20
14	63	17.8	449	1	US-08-102-942A-4
15	63	17.8	449	1	US-08-102-942A-6
16	63	17.8	2703	1	US-08-185-432-19
17	62	17.6	360	2	US-08-531-927B-2
18	61	17.3	423	1	US-08-523-376-3
19	59.5	16.9	220	4	PCF-US95-03236-36
20	59	16.7	260	3	US-08-558-135-7
21	59	16.7	432	2	US-08-933-750C-47
22	59	16.6	432	3	US-09-234-613-47
23	58.5	16.6	517	1	US-08-190-802A-30
24	58	16.4	362	1	US-08-437-027-21
25	58	16.4	363	3	US-08-651-999A-4
26	58	16.4	429	1	US-08-234-783-4
27	58	16.4	429	1	US-08-456-907-4
28	58	16.4	429	4	PCF-US95-05523-4

29	58	16.4	788	2	US-08-918-914-4	Sequence 4, Appl1
30	58	16.4	795	1	US-07-716-827C-5	Sequence 5, Appl1
31	58	16.4	1350	2	US-08-319-866-9	Sequence 9, Appl1
32	57	16.1	450	2	US-08-665-037-2	Sequence 2, Appl1
33	57	16.1	450	2	US-08-666-067-2	Sequence 2, Appl1
34	57	16.1	450	2	US-08-732-870-2	Sequence 2, Appl1
35	57	16.1	930	3	US-09-283-743-2	Sequence 2, Appl1
36	57	16.1	1805	1	US-07-853-913-2	Sequence 2, Appl1
37	56.5	16.0	2441	1	US-08-194-468-2	Sequence 2, Appl1
38	56.5	16.0	2441	3	US-08-961-739-2	Sequence 2, Appl1
39	56	15.9	55	1	US-08-469-427A-3	Sequence 3, Appl1
40	56	15.9	55	2	US-08-609-443B-3	Sequence 3, Appl1
41	56	15.9	55	2	US-08-569-063C-3	Sequence 3, Appl1
42	56	15.9	77	2	US-08-570-227A-11	Sequence 11, Appl1
43	56	15.9	188	1	US-08-469-427A-5	Sequence 5, Appl1
44	56	15.9	188	2	US-08-609-443B-5	Sequence 5, Appl1
45	56	15.9	188	2	US-08-569-063C-5	Sequence 5, Appl1

## ALIGNMENTS

```
RESULT 1
US-07-955-905A-2
; Sequence 2, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: Linear
; MOLECULE TYPE: protein
US-07-955-905A-2

Query Match 33.6%; Score 118.5; DB 1; Length 566;
Best Local Similarity 41.2%; Pred No. 66-06; Indels 9; Gaps 4;
Matches 28; Conservative 14; Mismatches 17;

QY 1 KRDPQQRHYEDCRRRCQQE---EPROQYOCORRC---RROQRHGRGCDLINDQCGSGR 54
:||||:|||||:|||||:|:|||||:|||||:|:|
Db 35 ERDPQQRHYEDCRRRCQQE---EPROQYOCORRC---RROQRHGRGCDLINDQCGSGR 91
QY 55 YEEGEKQ 62
Db 92 COEQOQOQ 99

RESULT 2
US-07-955-905A-22
; Sequence 22, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/955,905A  
FILING DATE: 21-JAN-1993  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 566 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORGANISM: Theobroma cacao  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..566  
OTHER INFORMATION: /note= "67 kD Precursor Protein"  
US-07-955-905A-22

Query Match 33.6%; Score 118.5; DB 1; Length 566;  
Best Local Similarity 41.2%; Pred. No. 6e-06; Indels 9; Gaps 4;  
Matches 28; Conservative 14; Mismatches 17;

QY 1 KRDPOREYEDRCRCEQ--EPROYOQCRRC---REORQHGRGGLINPGRSGR 54  
DB 35 EKDPRQ-QYEGQCRCESEATEREGCEQRCERKYNQQRQER--ELQNYOQQCQR 91

QY 55 YEEGEKQ 62  
DB 92 COEQOQOQ 99

RESULT 3  
US-07-955-905A-23  
Sequence 23, Application US/07955905A  
Patent No. 5770433  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND  
TITLE OF INVENTION: PRECURSOR  
NUMBER OF SEQUENCES: 28  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/955,905A  
FILING DATE: 21-JAN-1993  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 587 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Gossypium hirsutum  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..587  
OTHER INFORMATION: /note= "Vicilin from G. hirsutum"  
US-07-955-905A-23

Query Match 31.6%; Score 111.5; DB 1; Length 587;  
Best Local Similarity 52.4%; Pred. No. 3.8e-05;  
Matches 22; Conservative 8; Mismatches 7; Indels 5; Gaps 2;

QY 3 DPOQREYEDRCRCEQEPROYOQCRRC---REORQHGR 40

DB 81 DPQR-YEECQCECRQEEBRQPCQCFIKRFEQOQQOQSR 121

RESULT 4  
US-08-056-200-94  
Sequence 94, Application US/08056200  
Patent No. 5616500  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fredrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054,001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1898 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-056-200-94

Query Match 20.3%; Score 71.5; DB 1; Length 1898;  
Best Local Similarity 37.3%; Pred. No. 4.1;  
Matches 22; Conservative 10; Mismatches 24; Indels 3; Gaps 2;

QY 1 KRDPOREYEDRCRCEQEPROYOQCRCEQ--RQHGSGGLINPGRSGRYEE 57  
DB 1723 ERDRKFRFEQQLRQGRQQLRSQ--ESDRKFRFEQQLRQGRQQLRPQQRGKRYMEE 1780

RESULT 5  
US-08-800-644-94  
Sequence 94, Application US/08800644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb, Martens, Olson & Bear

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

	Query Match	Score	DB	Length
Best Local Similarity	20.3%	71.5	2	1898
Matches	22	Conservative	10	Mismatches
				24
				Indels
				3
				Gaps
				2

RESULT 6  
US-08-728-323A-2  
Sequence 2, Application US/08728323A  
Patent No. 5948676  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
TITLE OF INVENTION: Encoding Same And Uses Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728, 323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

```

? NAME: White, John P.
? REGISTRATION NUMBER: 28, 678
? REFERENCE/DOCKET NUMBER: 05/5,5268./JPM/MSC/SKRS
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-278-0400
? TELEFAX: 212-391-0525
? INFORMATION FOR SEQ. ID NO.: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1162 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
US-08-728-323A-2

```

Query Match	19.0%	Score 67	DB 2	Length 1162
Best Local Similarity	40.5%	Pred. No. 7.9		
Matches	15	Conservative	10	Mismatches 10; Indels 2; Gaps 1
QY	1	KRDQQRVEYEDCRRRCRCEQDEPRQQTQCCQRRCRCEQQRQ	37	
Db	566	QREPRQQRKEPQ--QREPRQQRKEPQQRKEPQQRKEPQQRQ	600	

```

1      RESULT
2      7
3      PCT-US95-03236-21
4      : Sequence 21, Application PC/TUS9503236
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: University of Southern California
9      :
10     : TITLE OF INVENTION: Methods to Diagnose and Treat HIV-1
11     : TITLE OF INVENTION: Infection
12     :
13     : NUMBER OF SEQUENCES: 66
14     :
15     : CORRESPONDENCE ADDRESS:
16     :
17     : ADDRESSEE: Campbell and Flores
18     : STREET: 4370 La Jolla Village Drive, Suite 700
19     : CITY: San Diego
20     : STATE: California
21     : COUNTRY: USA
22     : ZIP: 92122
23     :
24     : COMPUTER READABLE FORM:
25     :
26     : MEDIUM TYPE: Floppy disk
27     :
28     : COMPUTER: IBM PC compatible
29     : OPERATING SYSTEM: PC-DOS/MS-DOS
30     : SOFTWARE: PatentIn Release #1.0, Version #1.25
31     :
32     : CURRENT APPLICATION DATA:
33     :
34     : APPLICATION NUMBER: PCT/US95/03236
35     : FILING DATE: 13-MAR-1995
36     :
37     : CLASSIFICATION:
38     :
39     : ATTORNEY/AGENT INFORMATION:
40     :
41     : NAME: Imbira, Richard J.
42     : REGISTRATION NUMBER: 37, 643
43     : REFERENCE/DOCKET NUMBER: PE-SI 1394
44     :
45     : TELECOMMUNICATION INFORMATION:
46     :
47     : TELEPHONE: (619) 535-9001
48     : TELEFAX: (619) 535-8949
49     :
50     : INFORMATION FOR SEQ ID NO.: 21:
51     :
52     : SEQUENCE CHARACTERISTICS:
53     : LENGTH: 614 amino acids
54     : TYPE: amino acid
55     :
56     : TOPOLOGY: linear
57     :
58     : MOLECULE TYPE: protein
59     :
60     : CTT-US95-03236-21

```

Query Match Similarity 18.7%; Score 66; DB 4; Length 614;  
 Best Local Similarity 26.5%; Pred. No. 5.3;  
 Matches 18; Conservative 15; Mismatches 25; Indels 10; Gaps 2;

Db	440	RRRSRSRDKKEERRRSRERSKDKDRDRKRSSRSREARRERERKRELLRGGGDMAESEA	499
Oy	51	GSGRYEEG	58







```

;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-102-942A-6
    
```

```

Query Match          17.8%; Score 63; DB 1; Length 449;
Best local Similarity 25.0%; Pred. No. 8.2;
Matches 15; Conservative 14; Mismatches 25; Indels 6; Gaps 1;
    
```

```

QY      4 PQGREYEDCRRRCEDQEPPOQYCCOR-----RCREQQRQHGGRGDLINPQRGSGGRYEE 57
      | |:::| | | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db      352 PYQDFKDCERFRSRDQLKRRHQRHTGVVPFOCKTCQRRKFSRSDHLKTHRTHTGCTSE 411
    
```

Search completed: March 1, 2001, 15:49:36  
 Job time: 381 sec

